Investigations on the Distribution of Days Required to Attain Cumulative Precipitation Amounts

by

I. Herendi

Kumulatív csapadékösszegek eléréséhez szükséges napok eloszlásvizsgálata. A gyakorlati munkában gyakran szükséges az adott időn belül várható csapadék előrejelzése. A dolgozat egy módszert mutat be a statisztikai becslésre. Négy csapadékmérő állomás adatai alapján, az ún. csapadéktelítődési függvény bevezetésével a szerző a kumulatív csapadékösszegek eléréséhez szükséges napok eloszlásfüggvényeit vizsgálta. Az adatsorokra vonatkozó statisztikai próbák (függetlenség-, homogenitásvizsgálat) elvégzése után a *Gamma*-eloszlás paramétereinek kiszámítása, értékelése következett. Ez elegendő volt egyetlen állomáson (és kis környezetében) az előrejelzéshez. A területi általánosításhoz át kellett térni a *Pearson-III* típusú eloszlás alkalmazására. Így a módszer alkalmas lett nagyobb földrajzi egységekre vonatkozó statisztikai becslésekre.

Verteilungsuntersuchang der zum Erreichen von kumulativen Niederschlagsmengen erforderlichen Tage. In der praktischen Arbeit taucht oft die Notwendingkeit der Vorhersage eines innerhalb einer gegebenen Zeitdauer herabfallenden Niederschlagsmenge auf. In der vorliegenden Arbeit wird eine Methode zur statistischen Schätzung beschrieben. Auf Grund der Angaben von vier Niederschlagsstationen und mit der Einführung der sog. Funktion der Niederschlagssättigung werden vom Verfasser die Niederschlagsverteilungsfunktionen der zum Erreichen von kumulativen Niederschlagsmengen erforderlichen Tage untersucht. Nach der Ausführung der statistischen Proben bezüglich der Datenreihen (Untersuchung der Unabhängigkeit und Homogeneität) wurde die Errechnung der Parameter der *Gamma*-Verteilung, sowie ihre Auswertung vorgenommen. Dies war genegend auf bloss einer Station (und ihrer Umgebung) die Vorhersage ausführen zu können. Zur territorialen Verallgemeinerung musste auf die Anwendung der Verteilung vom Typ *Pearson-III* übergegangen werden. In dieser Weise wurde die Methode auch zu statistischen Schätzungen grösserer geographischen Einheiten geeignet.

In the practical work the forecasting of the precipitation to be expected within a given time is often required. In the present paper a method of the statistical estimation is described. On the basis of the precipitation data measured by four stations and by introducing the so-called function of precipitation-saturation the author gives an analysis of the distribution-functions of the days required for attaining the cumulative precipitation amounts. After the statistical checking of the data-series (analysis of independence and of homogeneity) the computation, evaluation of the parameters of the *Gamma*-distribution was carried out. This was sufficient for making the forecast for a single station (and its close environment). For the areal generalization the distribution Type *Pearson-III* had to be applied. Thus the method became suitable for statistical estimations concerning larger geographical units.

1. Introduction

Both the meteorologist and the geographer, when analysing complex processes, give particular attention to the precipitation from among the different parameters.

In many cases it would be desirable to know in advance the precipitation amount to be expected for a given area at a given fixed time.

This practical requirement gave rise to the idea of elaborating a method enabling the specialist to determine, within certain limits of exactitude, the probable precipitation amount at a given geographical point within a certain time, or to state the probable precipitation income within a concrete limit of probability.

It was envisaged also to make the method to be elaborated applicable also for the general evaluation of well measurable geographical parameters (that can be expressed by data series). The aim of the method was also to give assistance in the analysis and, later, in the synthesis, when dealing with complex processes.

2. Objectives

Anwers are to be found to the following questions:

1. Which is the percental probability of a precipitation amount exceeding, within a fixed time, a value determined in advance?
2. Which precipitation amount can be expected with a given probability within a given period?
3. After how many days is a given precipitation amount to be expected with a given probability?
4. How can generalization in time be made for an optional day of the year but for the same place?
5. How is the areal generalization to be solved?

3. Selection of the observation places

Serious problems arose when selecting the stations disposing of suitable data series. For the present there are more than 700 precipitation observation points in Hungary but comparatively few of them dispose of time series of the completeness and length required for our purposes.

From among the stations Budapest, Szeged, Keszthely and Eger proved to be the most appropriate. By the more than 100 years old daily data series of the above station a good representation of the climatic peculiarities of our geographic regions can be obtained facilitating also the areal generalization (*s. Fig. 1*).

4. Mathematical bases

Our task is now to determine the probability of the excession of a certain precipitation amount at one and later at several points of the space, and also the connection

$$y = f(t, p(t))$$

of the above and the time [3].

The trend of precipitation values is the result of a stochastic process. Under a stochastic process the one parameter assembly of probability variables ξ_t is meant, where the parameter t runs through a — generally — T set and is denoted with

$$\xi_t,\ t \in T$$

If the variable ω (elementary occurrence) of the essentially two-variable-function $\xi_t = \xi_t(\omega)$ (t and ω) is fixed and t runs through set T, than a real function, a realiza

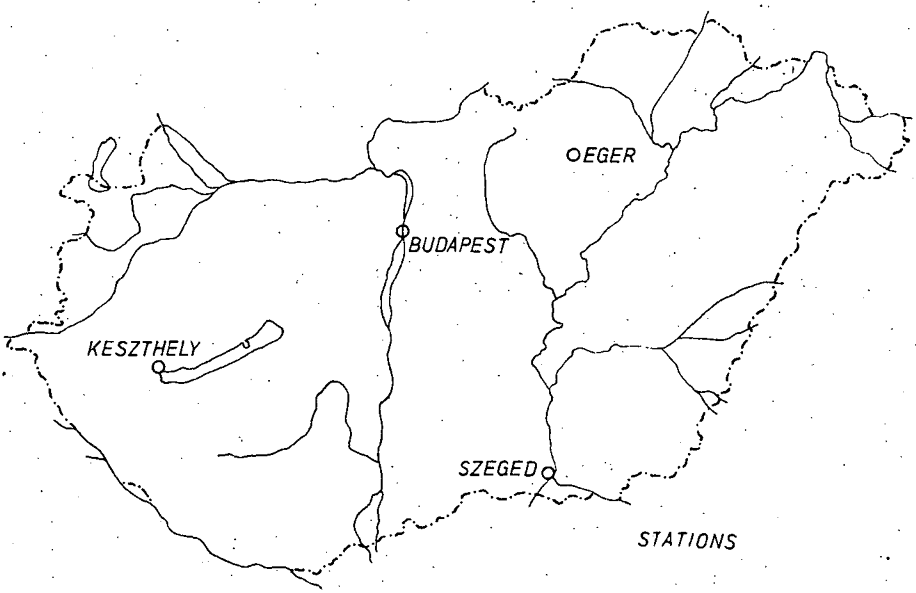

Fig. 1. The situation of the investigated stations
1. ábra A vizsgálatba bevont állomások elhelyezkedése

tion of the stochastic process will take place. One realization will characterize a concrete course of the process.

The measured values of the precipitation are concrete values of the probability-variable ξ_t.

The examined phenomenon, the precipitation, can be solved with discrete data series. From a different aspect our probability-variable is limited from below (it cannot assume negative values).

The distribution of the probability variable is unknown. As to its properties only inferences can be made from the finite data series called "model" permitting also computations, statistical evaluations concerning the attributes of the basic set.

In order to make inferences from our models as to the distribution of the basic set the models have to meet two basic conditions: they must be (with a good approximation) independent and homogeneous.

A most utilizable attribute of the precipitation data is that their function of distribution can be readily evaluated on the basis of models consisting of discrete data. This facilitates also the use and application of the method, to be presented by us.

Under the distribution-function of a probability variable the following is meant "per definitionem":

$$F(x) = P(\xi < x), \quad -\infty < x < +\infty$$

The basic features of the distribution functions are the following:
monotonously increasing:

$$F(x_1) < F(x_2), \quad \text{if} \quad x_1 < x_2$$

continuous from the left:

$$\lim_{x \to x_0} F(x) = F(x_0)$$

and also:

$$\lim_{x \to -\infty} F(x) = 0$$

and

$$\lim_{x \to +\infty} F(x) = 1$$

The distribution function $F(x)$ describing the distribution of the basic set is estimated with the empirical distribution function $F_n(x)$ computed from the model, if in the case of model $\xi_1, \xi_2, \ldots, \xi_n$

$$F_n(x) = \begin{cases} 0 & \text{if} \quad x \leqq \xi_1 \\ \dfrac{k}{n} & \text{if} \quad \xi_{k-1} \leqq x < \xi_n \\ 1 & \text{if} \quad \xi_n \leqq x \end{cases}$$

The distribution functions are to be ranged among the most frequently occurring classes of distribution functions. Thus the originally infinite set can be considerably narrowed down to a subset of the utilized types of distribution. The reason of the occurrence of the distribution types in nature is supported by several theorems of probability calculation — for instance the central limit distribution theorems.

On the basis of a finite model the determination of the type of $F(x)$ is facilitated when drawing the empirical distribution function of the model, i.e. $F_n(x)$ and comparing the curve with those of the more known and more manageable distributions. Selecting from among them the most appropriate one we have to determine, by the aid of the model elements, the form and parameters of the function. The empirical and theoretical distribution functions can be brought even more exactly together by the use of the methods of mathematical adaptation investigations.

5.1 The applied methods

The initial series, used for the computations, were the 100 years daily precipitation series of the selected stations. From these were produced the values determining the so-called "function of precipitation saturation". This is a step-function and such ones were attached to every fifth day of the year and to 5 precipitation amounts determined in advance. Thus the concrete values of these step-functions will yield, for the initial days being in a distance of five days one from the other, the number of the days within which 20. 30, 50, 70, 100 mm will fall down. (E.g.; the first initial day is the lst of January, the second one the 6th, the third one the 11 January etc.)

So at each station five figures were attached to every fifth day of the year. This means also that the 36 500 data collected from each station were transformed to further 36 500 data.

From the series produced by transformation we selected and ordered the number of excess days (or saturation values) of 100 different years but belonging to the same initial days and to the same cumulative amounts.

For these ordered series some simple statistical parameters too, have been computed which were helpful in our further work.

5.2 The investigation of independence

In the foregoings suppositions were made concerning our models. Their fulfilment is now to be checked.

The independence of the model elements means the following:

The probability variables ξ and η are called independent, if in the case of optional figures $a \leqq \xi \leqq b$ and $c \leqq \eta \leqq d$ the following equality is accomplished:

$$P(a \leqq \xi \leqq b,\ c \leqq \eta \leqq d) = P(a \leqq \xi \leqq b).\ P(c \leqq \eta \leqq d).$$

This is equivalent with the condition according to which:

$$H(x, y) = F(x) \cdot G(y),$$

where on the left side we have the joint distribution function of the probability variables ξ and η, and while $F(x)$ and $G(y)$, appearing on the right side, are the distribution functions of the random variables ξ and η.

The independence investigations were carried out with the method of *Wald-Wolfowitz.* [5]

According to the theorem functioning as the basis of the control:

In the case of an independent sample of a large number of elements, if the elements arise from an identical distribution, the sum,

$$R = \sum_{i=1}^{n-1} \xi_i \cdot \xi_{i+1} + \xi_1 \cdot \xi_n$$

formed from it will be, with a good approximation, of normal distribution.

Its expectable value is:

$$M(R) = \frac{S_1^2 - S_2}{n-1}.$$

Its variance

where n = the number of element of the sample

$\xi_i =$ the i-eth element of the sample arranged in the order of observation

$$S_i = \sum_{j=1}^{n} \xi_j^i.$$

For the control of the independence the distribution of $|R|$ is employed, and after that the required percentual probabilities are computed on the basis of the formula

$$p\% = 2 \cdot (100 - f(X))$$

where

$$|Y = \frac{R - M(R)}{D(R)}.$$

The "sample of large number of element" requires that $n \geqq 30$. Since in our case $n = 100$, this condition is fulfilled.

5.3 Investigation of homogeneity

In the foregoings reference was made to our hypothesis of the evenness of our samples. This too, is a most important criterion, since, although in our precipitation data series and at least on the examined places, no important change of climate could be proved, but e.g., a change of the surroundings of the observation place or of the type of the instrument could lead to inhomogeneities in our data. This investigation may appear as an important aspect also in evaluating other parameters.

The examination of homogeneity is based on the theorem of *Smirnov*. In this theorem it is stated that if two samples (with the element-nombers k and j) originate from a basic set of identical distribution, and they are independent from each another, so, when looking for the maximum absolute value of the differences ($d_{k,j}$) of the differences between the empirical distribution functions formed from them, the

$$Z = \sqrt{\frac{k+j}{k \cdot j}} \cdot d_{k,j}$$

product is, with a good approximation, the probability variable of *Kolmogorov*, in the case of $k, j > 30$.

The samples examined by us have been disjoined into such part-samples, and these investigations have been carried out.

The $L(Z)$ probability of the obtained values Z can be found in the standard tabulation of the *Kolmogorov*-distribution. The percentile probabilities, characteristic for the homogeneity, are to be computed from the formula

$$p\% = 100(1-L(Z))$$

5.4 The use of the distribution functions

On the basis of the initial concepts, in meteorology and water economics we choose the *Gamma*-distribution with the distribution function

$$F(x) = \frac{\lambda^k}{\Gamma(k)} \int_{x_0}^{x} t^{k-1} \cdot e^{-(t-x_0)}\, \mathrm{dt}.$$

Hypothetically it is supposed that the sample originates from the basic set of the theoretical distribution $F(x)$.

The curves of $F(x)$ and $F_n(x)$ will not correspond in all points but the deviations and their place within the domain of interpretation will be characteristical of the fitting.

Our initial hypothesis was that the two distributions are identical. This is the so-called zero-hypothesis. The question to be solved is whether or not this hypothesis is correct at the given significance level.

The hypothesis is to be rejected if in some part of the domain of interpretation the deviations are too large and unidirectional. If the hypothesis can be kapt, the deviations can be considered as random fluctuations of the sample.

The distribution-function of the *Gamma*-distribution has two parameters: the formal parameter "k" and the scale-parameter "λ". In its general form x_0 *too,* is a parameter, but in our case $x_0=0$, since that number of days is required to the precipitation of zero-mm.

For the estimation of the parameters "k" and "*lambda*" of the distribution function two methods can be used:

a) in the case of the so-called maximum likelihood:

$$k = \frac{1}{4 \cdot A}\left(1 + \sqrt{1 + \frac{4 \cdot A}{3}}\right)$$

$$\lambda = \frac{k}{\bar{x}}$$

where:

$$A = \ln \bar{x}$$

$$\bar{x} = \sum_{i=1} x_i/n$$

b) in the case of the method of moments:

$$k = \frac{m_1^2}{m_2^*}$$

$$\lambda = \frac{m_1}{m_2^*}$$

where m_1 is the value to be expected, while m_2^* is the second central moment or, as it is more often called, the variance. (It is customary to name the expected value also "first moment": hence the denomination of the method [4].

The question arises why our random variable (describing the precipitation data and being of continuous distribution) is treated discretely. This can be explained partly by the practical execution of the sampling, and partly by the circumstance that in the case of certain conditions the theoretical continuous distribution-function can be estimated very well with the discrete distribution.

6. Extension to any arbitrary day of the year and to arbitrary cumulative precipitation amounts

The original aim of the investigations was to find certain parameters for the statistical forecasting of the precipitation, and to extend these parameters in time and space to larger geographical units.

The processing has been made for the series of Budapest, Keszthely, Szeged and Eger. The transformation of the basic data, i.e. the production of the saturation values were carried out. After that, the produced samples were examined as to whether they satisfy the required, and from the aspect of the processing indispensable, conditions. The samples were found homogeneous, independent and well fitting (*Fig. 2; Tables 1—4*).

The analysis of the parameters (characteristic of the samples) has proved that the aspects fixed on the basis of preliminary considerations are good and suitable for the generalization (*Fig. 3*).

The results are already applicable to each fifth starting day and to five precipitation limits.

Now it is to be examined how the previous considerations could be generalized for an arbitrary day of the year and an arbitrary precipitation limit at a station or within its smaller surroundings.

Since the $[(n-1):5+1]$-eth day of the year has been chosen arbitrarily by us it seems obvious to condense the initial days (fixed when computing the saturation times) by changing the operandus "5". With an analogous procedure arbitrary precipitation limits can be established also instead of 20, 30, 50, 70 and 100 mm. But that would have meant a considerable increase of the time required to the processing.

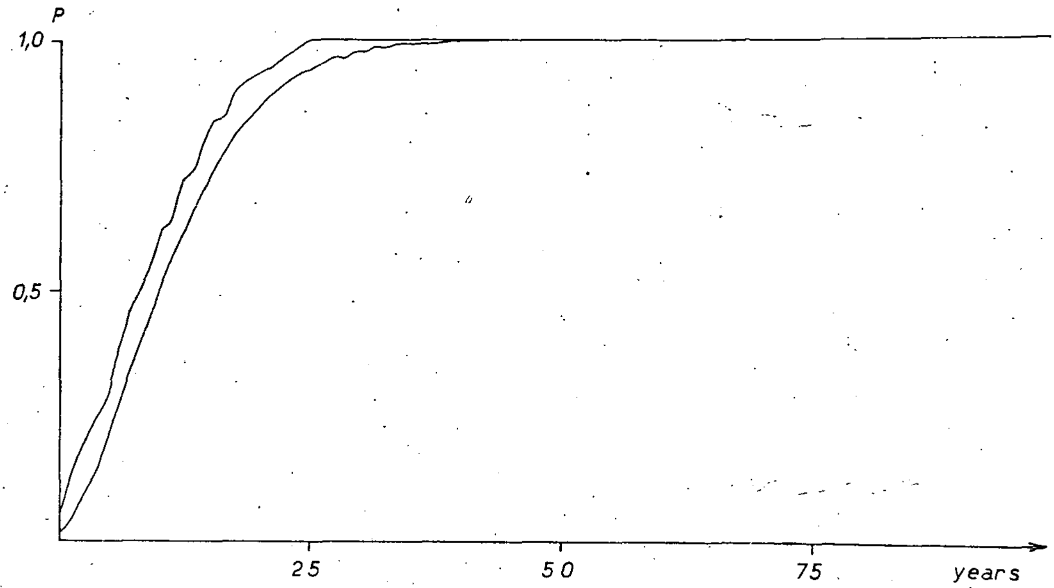

Fig. 2. Investigation of fitting. Budapest, lst starting day January 1. Fitting of distribution function belonging to the cumulative precipitation amount of 20 mm
2. ábra. Illeszkedésvizsgálat. Budapest, első kezdőnap január 1. 20 mm-es kumulatív csapadékösszeghez tartozó eloszlásfüggvény illeszkedése

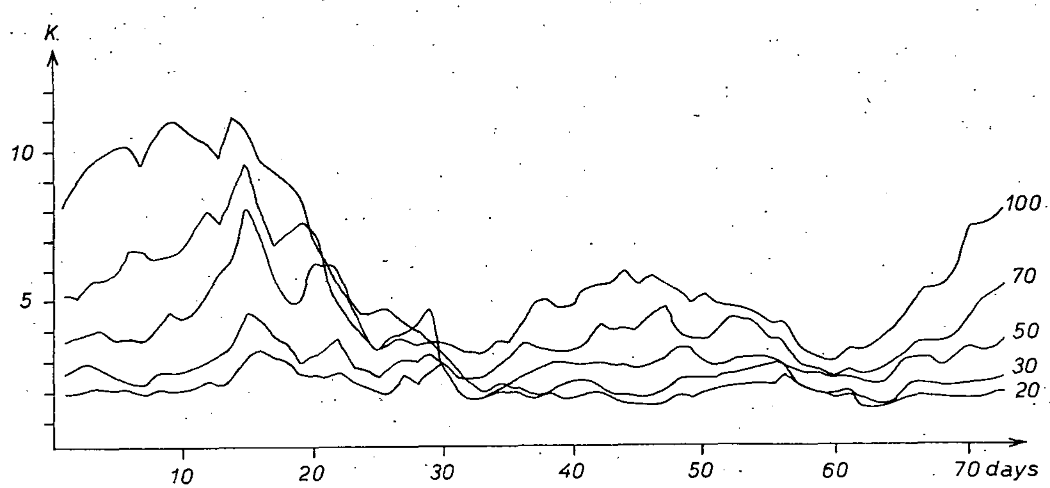

Fig. 3. Investigation of the parameters of the distributions. The „k" parameters of Budapest during the year, for 5 precipitation amounts
3. ábra. Az eloszlások paramétereinek vizsgálata. Budapest „k" paraméterei az év során, 5 csapadékösszeghez

Table 1

Investigation of fitting
Budapest, initial day 53, cumulative limit 20 mm empirical values

1.	0,03	0,05	0,10	0,11	0,12
2.	0,14	0,16	0,23	0,25	0,31
11.	0,32	0,34	0,43	0,48	0,53
16.	0,57	0,58	0,59	0,61	0,62
21.	0,65	0,67	0,69	0,70	0,73
26.	0,74	0,75	0,78	0,80	0,82
31.	0,84	0,85	0,86	0,88	0,89
36.	0,92	0,93	0,94	0,95	0,97
41.	0,98	0,99	1,00	1,00	1,00
46.	1,00	1,00	.	.	.
	.	.	.	.	.
91.	1,00	1,00	1,00	1,00	1,00
96.	1,00	1,00	1,00	1,00	1,00

Table 2

Investigation of fitting
Budapest, initial day 53. cumulative limit 20 mm computed values

1.	0,00	0,01	0,03	0,05	0,07
6.	0,09	0,13	0,16	0,19	0,22
11.	0,25	0,29	0,32	0,35	0,38
16.	0,41	0,44	0,47	0,50	0,53
21.	0,55	0,58	0,60	0,63	0,65
26.	0,67	0,69	0,71	0,73	0,74
31.	0,76	0,78	0,79	0,80	0,82
36.	0,83	0,84	0,85	0,86	0,87
41.	0,88	0,89	0,90	0,91	0,92
46.	0,92	0,93	0,93	0,94	0,94
51.	0,94	0,95	0,95	0,95	0,96
56.	0,96	0,96	0,96	0,97	0,97
61.	0,97	0,97	0,98	0,98	0,98
66.	0,98	0,98	0,98	0,98	0,99
71.	0,99	0,99	0,99	0,99	0,99
76.	0,99	0,99	0,99	0,99	0,99
81.	0,99	0,99	0,99	0,99	1,00
86.	1,00	1,00	1,00	1,00	1,00
91.	1,00	1,00	1,00	1,00	1,00
96.	1,00	1,00	1,00	1,00	1,00

Table 3

Investigation of homogeneity (Budapest)

Initial day	*Cumulative limit mm*	*Homogeneity %*
1	20	100,00
1	70	99,99
1	100	100,00
53	20	100,00
53	70	100,00
53	100	100,00

Table 4

Investigation of independence
(Budapest)

Initial day	Cumulative limit mm	Independence %
1.	20	82,58
1.	30	100,00
1.	50	96,80
1.	70	100,00
1.	100	95,22
53.	20	60,30
53.	30	79,48
53.	50	82,58
53.	70	93,62
53.	100	100,00

Table 5

Correlation coefficients (Szeged—Eger)
30. initial day
Correlation matrix, Szeged

	20 mm	30 mm	50 mm	70 mm	100 mm
20 mm	1,00	0,83	0,70	0,54	0,48
30 mm	0,85	1,00	0,81	0,63	0,57
50 mm	0,70	0,81	1,00	0,76	0,70
70 mm	0,54	0,63	0,76	1,00	0,86
100 mm	0,48	0,57	0,70	0,86	1,00

Correlation matrix, Eger

	20 mm	30 mm	50 mm	70 mm	100 mm
20 mm	1,00	0,85	0,65	0,60	0,53
30 mm	0,84	1,00	0,79	0,71	0,57
50 mm	0,66	0,79	1,00	0,92	0,78
70 mm	0,60	0,71	0,92	0,99	0,85
100 mm	0,53	0,57	0,78	0,85	1,00

Correlations between the corresponding limits (Szeged—Eger)

20 mm	0,39
30 mm	0,49
50 mm	0,55
70 mm	0,45
100 mm	0,35

In the case of a continuous parameter there is also a more serviceable method, but even then it must be considered whether it is worth while to apply a raster that is even more dense than a certain resolution.

On the basis of our experience it can be stated that two days, near to each other, e.g. two subsequent days, may be in close connection from the aspect of the precipitation. It is comparatively easy — but supportable also with computations — to suppose that there is a linear connection between the subsequent days and their respective saturation-, cumulation-values, or more exactly: between the distributions of the saturation values belonging to the days following each other rather densely.

So the linear connection has been supposed. To prove this, a new statistical method the correlation computation was applied (*Tables 3 and 5*).

In the examined cases comparatively high values with positive signs were obtained, and on the basis of the significance investigations it may be stated that according to the linear character and density of the connections the extension for time and quantities can be solved by means of the linear interpolation between the data. And this brings not only a gain in the time but facilitates also the practical application of the method. Thus if we are looking for a value for a given distribution in our tables as to an arbitrary initial day of the year or to arbitrary cumulative amounts, then we have to linearly interpolate between the corresponding data of two, already existing neighbouring distribution.

7. Three examples for the application of the method

Thus on the basis of the applied methods, considerations and computations we are already in a position to give the answer to the questions relating only to *one* measuring point from among those enumerated in our program. Since the measuring points represent as a rule only restricted surroundings, the reliability of the answers will rapidly decrease in proportion with our moving away from the measuring point.

In the possession of the respective tables the following questions may be answered:

1. *What is the percentage of probability of a given precipitation amount within a given time?*

For that the number of the respective starting day and the respective limit must be looked for in the table.

Examining the ordered sample elements we will arrive to the value of the fixed time (i.e. days). The place where it will be found is the position index (POZ) and that shows the required probability on the basis of the following formula:

$$p\% = \frac{100 \cdot \mathrm{POZ}}{n}$$

where n means the number of years included in the elaboration.

2. *In how many days can be expected a given precipitation amount with a given probability?*

First of all the serial number of the initial day is to be determined.

The n-th initial day is the serial day

$$(n-1):5+1$$

of the calendar year.

In the table of the empirical distributions one has to find the required starting day with the required precipitation limit (cumulative precipitation amount). Such values are to be found e.g. in *Table 6*.

After that we will compute from the following proportion:

$$\frac{100}{\text{number of the sample elements}} = \frac{\text{given probability } p}{\text{serial number of the sample element}}$$

that is:

$$100 : n = p : \text{POZ}$$

From this:

$$\text{POZ} = \frac{n \cdot p}{100}$$

and so the required value will be the number figuring at this place.

Our investigations were carried out for the data series representing $n=100$ years. Thus $n=100$. So in the tables at our disposal the measuring number of the given probability will be the serial number of the place where the number of the days required to the accumulation can be found.

Some difficulties arise merely from the correct setting of the initial day because in our tables only every fifth one has been fixed. However, as it has been shown, that can be helped by linear interpolation.

3. *A precipitation amount of how many millimeters can be expected within a given period and with a given probability?*

On the basis of the formula the actual starting day is to be looked for and after that we have to compute — if it is not given so — the number of days corresponding to the period and also the position corresponding to the given probability.

On the basis of that we select that sample from among the five ordered ones (precipitation limits of 20, 30, 50, 70 and 100 mm) in which at the so determined place the value, falling next to the obtained period, is found. That will be the required value.

8. On areal generalization

On the basis of the above considerations and computations the extension to an arbitrary day of the year can be considered as solved. A much more complicated problem is presented by the areal generalization.

By reason of the obtained results the application of the distribution functions seems to offer favourable possibilities.

The *Gamma*-distribution is determined by the parameters "*k*" and "*Lambda*". For producing them the expected value m_1 and the variance m_2 are needed, from which

$$\lambda = \frac{m_1}{m_2^*}$$

and

$$k = \frac{m_1^2}{m_2^*}$$

The computations have been carried out for four stations: Transdanubia is well represented by *Keszthely*, the Hungarian Lowland by *Szeged*, and the Northern Central Mountains (Északi Középhegység) by *Eger*.

The processing of the series of *Budapest* seemed to be serviceable because that station is situated at the meeting place of our mean geographical regions, and, on the other hand, almost all climatological-meteorological parameters measured there can be found here in well ordered and from many aspects in detail processed form. This is most of all from the aspect of the comparations of importance and use.

In addition to that we dispose of longer or shorter series concerning about 800 points in Hungary, or in other words: the monthly mean values of our presently investigated parameter (precipitation) of 800 geographical points are known to us.

As it can be seen, four stations are known to us in detail, while a certain smaller surrounding area of them — where the series of the measuring point can still be considered as representant — can be well characterized by the aid of our method.

Besides of the above we have at our disposal the nearly 800 measuring points but the fact is that either the series of these stations does not meet the statistical requirements of the sampling, or, apart from the monthly mean values we have possibly no other information whatever about them.

By this antagonism the following problems arise:

1. *How dense must be the network in order to allow a generalization of our method for our main regions and the whole country?*
2. *If at a given measuring station only the monthly means are known, how can the method be applied there and its smaller surrounding areas?*

8.1 How many stations are required at the minimum for the generalization for the whole country?

In order to give answer to the above question a detailed investigation of the basic data of the four selected stations, the series transformed from them, the different parameters, the distributions and the results obtained by the examination of fittings has been carried out (*Fig. 3, Table 6*).

Surprisingly many similarities were found during these investigations. We were mainly interested in the "attitude" of the parameters (describing these distributions) in the case of a common initial day or in that of a fixed precipitation limit.

The two parameters of the *Gamma*-distribution seemingly do not show any important variance: of course only in the case if out of the parameter-series of the four stations only the overlapping, values, i.e. those in identical position, are compared. The deviation is even in percentage not significant in the case of the compared data. The substantial conformity of the distribution parameters brought us to the idea of possibly construct parameter-series (valid for the entire territory of the country) from the values of "*k*" and "*Lambda*" previously computed for the investigated stations?

The checking has been made for five data series: for Budapest, Szeged, Eger, Keszthely, and for the arithmetical mean of the respective data of these four stations. In this way it was envisaged to represent Hungary by the fifth fictive station. The elements of this series was denoted by P_M.

In a formula the computing of the *i*-eth such parameter is the following:

$$P_M^{(i)} = \frac{P_{\mathrm{BP}}^{(i)} + P_{\mathrm{SZ}}^{(i)} + P_{\mathrm{K}}^{(i)} + P_{\mathrm{E}}^{(i)}}{4}.$$

By the checking computations, requiring a vast integration work, it has been proved that the indicated way means indeed the correct solution. The reproduced data series of the stations showed a good correspondence with the original ones and they cover each other very well. Thus, within a confidence- interval of some days, the individual stations may be substituted even with each other.

Table 6

Ordered saturation values, Szeged, 1. initial day

	20 mm	30 mm	50 mm	70 mm	100 mm
1.	4	4	12	19	28
2.	4	8	12	34	41
3.	4	8	12	33	41
4.	5	8	20	34	42
5.	5	8	23	35	45
6.	6	9	24	37	49
7.	7	10	25	37	50
8.	7	10	27	40	57
9.	7	11	27	40	57
10.	7	11	27	40	57
11.	7	12	29	40	57
12.	8	12	31	42	60
13.	8	14	31	42	61
14.	8	15	31	42	63
15.	8	15	32	43	63
16.	10	16	32	43	64
17.	10	16	33	43	64
18.	10	17	34	44	65
19.	10	17	34	44	65
20.	11	17	34	44	65
21.	11	17	35	45	66
22.	11	18	35	46	67
23.	11	18	36	47	67
24.	12	19	36	48	67
25.	12	19	36	49	69
26.	12	19	37	49	70
27.	13	19	37	50	70
28.	13	20	38	51	71
29.	13	21	38	51	73
30.	14	22	39	51	74
31.	14	22	39	51	76
32.	14	23	39	53	76
33.	14	24	39	53	79
34.	14	24	40	53	79
35.	15	24	41	53	81
36.	15	24	41	54	81
37.	15	24	41	59	83
38.	17	25	42	60	83
39.	17	25	42	60	85
40.	18	25	43	60	87
41.	18	25	43	60	87
42.	18	26	43	61	88
43.	18	26	44	62	89
44.	18	26	44	63	91
45.	18	27	44	63	92
46.	19	27	44	63	92
47.	19	27	45	66	93
48.	19	27	46	67	96
49.	20	30	47	69	97
50.	21	30	47	71	97
51.	21	30	48	71	98
52.	21	31	49	71	99
53.	22	32	50	72	99
54.	22	32	51	73	99
55.	23	33	52	75	99
56.	24	33	53	75	100
57.	24	33	53	77	101
58.	25	33	54	77	102

	20 mm	30 mm	50 mm	70 mm	100 mm
59.	25	33	55	78	102
60.	25	35	55	79	103
61.	26	35	55	79	104
62.	27	35	56	81	105
63.	27	35	57	82	105
64.	28	37	58	83	107
65.	28	37	58	85	107
66.	29	38	60	85	407
67.	29	40	61	86	108
68.	30	40	61	87	109
69.	30	40	62	87	110
70.	31	41	63	87	110
71.	31	41	68	88	112
72.	31	41	69	88	113
73.	32	44	71	89	113
74.	32	44	71	90	114
75.	32	45	73	93	115
76.	32	45	75	93	115
77.	33	47	76	93	116
78.	33	48	78	95	117
79.	34	48	79	96	118
80.	34	48	81	97	118
81.	34	50	83	97	118
82.	35	50	84	100	119
83.	35	52	85	100	119
84.	35	52	86	100	120
85.	39	55	86	101	124
86.	40	56	87	101	126
87.	42	65	89	102	128
88.	44	66	93	102	129
89.	65	67	95	106	131
90.	55	67	95	108	132
91.	55	69	97	111	133
92.	58	72	99	114	135
93.	59	74	99	118	135
94.	60	77	101	121	134
95.	64	78	102	123	135
96.	64	84	112	124	137
97.	72	90	115	125	144
98.	73	97	117	129	145
99.	76	107	122	132	150
100.	84	111	129	133	154

On account of the small areal variance the following considerations seem to be servicable both from the practical but even from the theoretical aspect:

1. The parameter-series of the detailedly investigated stations are to be considered as the series "k" and "λ" valid for the main geographical regions represented by the stations.
2. The values $P_M^{(i)}$, i.e. the mean values of "k" and "λ" are considered as characteristic distribution parameters valied for the whole country.

Thus we have determined the minimum station network density for our main regions, i.e. the distribution parameters required for the application of the method. However, it is also supposed that within smaller areas these investigated parameters may be possibly even more variable.

8.2 Territorial generalization with the knowledge of the monthly averages

To find an answer to this question is already considerably more difficult. It must be held in view that the mere knowledge of the monthly averages is not sufficient to the reproduction of the time series with the help of the parameters.

If the use of the *Gamma*-distribution is decided then a limit will be set up to the further computations by the requirement of the second central moment of the data in order to produce the parameters "k" and "λ". The variance is, however, not always accessible for us.

The following considerations were applied:

The parameters concerning the *Gamma*-distribution show, as it could be seen before, an extraordinarily small variance in the whole country. Even the data series of Transdanubia, having considerably more precipitation, and those of the Lowland (Keszthely — Eger) showing other characteristics, present only insignificant deviations. Thus on the basis of the preceding chapter stabile dataseries of "k" and "λ" were obtained satisfyingly characterizing the whole country.

Under point 5.4 and in connection with the distribution functions it was mentioned that several types of distribution functions occur in the practice and nature.

In the practice of hydrological work the distribution function of the type *Pearson-III* too, proved to be very good, in addition to the *Gamma*-distribution.

Accordingly, appropriately well elaborated tables are at our disposal in order to give assistance in the computations with it. Its greatest advantage is its generally more simple manageability in the practice, compared with the other types of distribution. In consequence of that, and in addition to its theoretical value, the practical specialists are first of all those who make a great use of it.

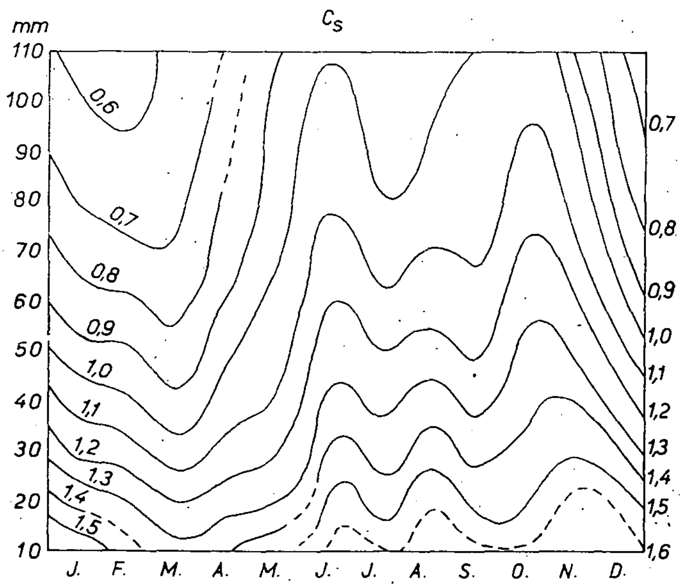

Fig. 4. Isoplethes of the values C_s of the distribution type Pearson-III
4. ábra. A Pearson-III eloszlás C_s értékeinek izoplétái

For our purposes the application of this type of distribution is essential because on the basis of the existing parameter-series one is enabled to easily switch over from the *Gamma*-distribution to the type *Pearson—III* which is rather similar to it.

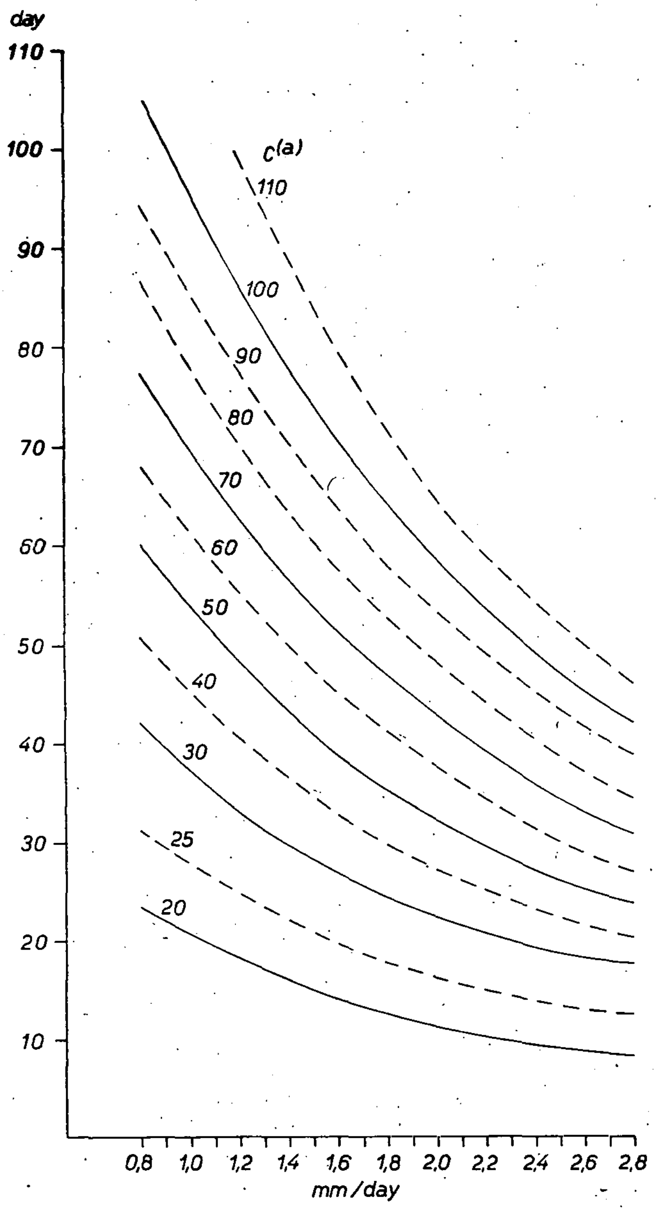

Fig. 5. The values $C^{(a)}$ of the distribution Pearson-III
5. ábra. A Pearson-III eloszlás $C^{(a)}$ értékei

The respective formulae are:

$$C_v = \frac{k}{m_1}$$

$$C_s = \frac{2 \cdot k}{m_1^3 \cdot C_v^3 \cdot \lambda^3} \cdot$$

C_v is the variance factor, and C_s the asymmetry-factor of the *Pearson—III* distribution.

With the knowledge of the parameters the computation method of the saturation value $t_p^{(a)}$ belonging to the limit "a" and probability "p" is the following:

$$t_p^{(a)} = [(\Phi(C_s, p)) \cdot C_v + 1] \cdot M,$$

where (C_s, p) is the value of the Pearson—*III* distribution function in the case of an asymmetry factor C_s and a selected transgression probability p. The values (C_s, p) can be found in the *Foster-Ribkin* standard table [7].

When analyzing the time trend of the parameters it appeared that their formation shows a marked yearly tendency, so that their values ave been monthly averaged for the purpose to clearly see the characteristic form of the yearly trend by filtering out the random variations.

As an example the monthly averages of the parameters M, C_v and C_s belonging to $C^{(a)} = 50$ mm have been elaborated in detail for the station Szeged.

From among the parameters C_v and C_s do not show within a month considerable variations at the stations representing the different geographical regions of the cou-

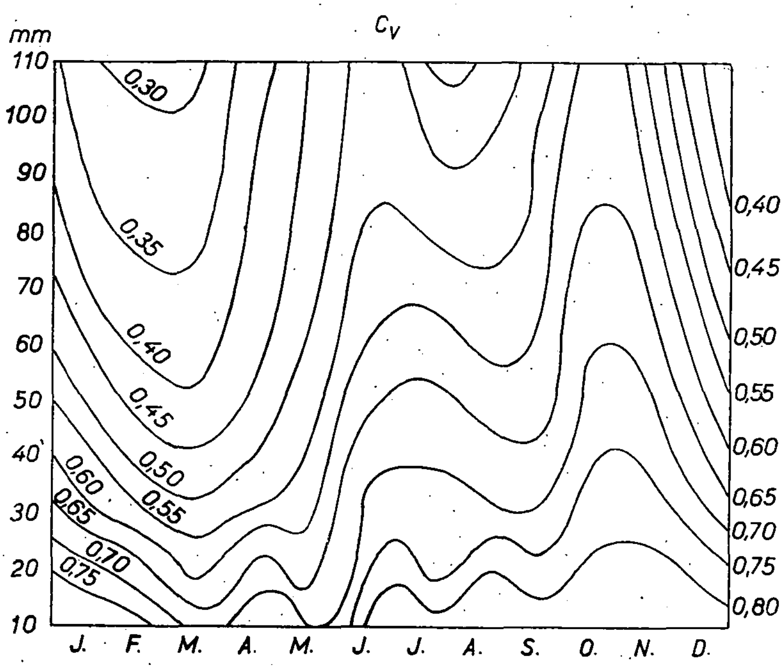

Fig. 6. Isoplethes of the values C_v of the Pearson-III distribution
6. ábra. A Pearson-III eloszlás C_v értékeinek izoplétái

try and so their arithmetical mean can be accepted as a characteristic standard value for the territory of the whole country. From the national averages of the parameters isoplethes were constructed (*Fig.-s 5 and 6*) from which not only the C_v and C_s values belonging to the selected precipitation amounts of 20, 30, 50, 70 and 100 mm but also those belonging to any intermediate precipitation amount $C^{(a)}$ can be found, because of their close connection with the amounts $C^{(a)}$.

The arithmetical mean M, on the other hand, shows considerable deviations according to the stations. However, the connection between M and K is obvious, where K is the mean precipitation amount of a given period after the time t_0, since the greater that average precipitation K the shorter time t will be required to reach the given cumulative precipitation amount $C^{(a)}$. For discovering the connection the most serviceable would be to know the average precipitation amount of periods of different durations after a certain given time datum. But such processings are not to our disposal, so only the known average precipitation amounts of the calendar months can be taken as a basis when analyzing the connection $M=f(K)$.

In our present investigations that the M values related to a given month N have been brought into connection with the average precipitation amounts of the months N, $N+(N+1)$, $N+(N+1)+(N+2)$. In order to eliminate the changing duration of the months, instead of the average precipitation amounts of K months only their part falling to 1 day have been taken into consideration, when carrying out the inves-

Table 7

Parameters of the distribution Pearson-III monthly
Eger, limit 30 mm

Month	Monthly average mm	Lambda	K	C_v	C_s
1.	30,70	0,07	2,22	0,69	1,34
2.	27,00	0,08	2,31	0,70	1,31
3.	22,40	0,13	2,99	0,59	1,15
4.	16,70	0,14	2,81	0,71	1,19
5.	16,20	0,17	2,45	0,56	1,27
6.	13,20	0,12	1,63	0,80	1,56
7.	16,20	0,12	2,00	0,72	1,41
8.	19,30	0,08	1,49	0,79	1,63
9.	22,50	0,08	1,83	0,75	1,47
10.	21,10	0,08	1,88	0,81	1,45
11.	21,20	0,07	1,43	0,80	1,67
12.	25,60	0,06	1,49	0,79	1,63

Keszthely, limit 20 mm

Month	Monthly average mm	Lambda	K	C_v	C_s
1.	25,90	0,08	2,26	0,72	1,33
2.	23,50	0,10	2,43	0,66	1,28
3.	20,80	0,12	2,52	0,63	1,25
4.	15,30	0,16	2,43	0,63	1,28
5.	13,50	0,15	2,07	0,71	1,39
6.	13,20	0,15	1,54	0,85	1,61
7.	14,00	0,15	2,08	0,68	1,38
8.	15,70	0,10	1,55	0,79	1,60
9.	16,40	0,10	1,72	0,74	1,52
10.	17,40	0,08	1,39	0,84	1,69
11.	22,30	0,08	1,88	0,76	1,45
12.	24,01	0,08	1,92	0,74	1,40

Budapest, limit 20 mm

Month	Monthly average mm	Lambda	K	C_v	C_s
1.	23,30	0,08	1,99	0,75	1,41
2.	23,20	0,09	2,11	0,69	1,37
3.	20,30	0,13	2,47	0,57	1,27
4.	15,70	0,15	2,37	0,65	1,29
5.	13,50	0,16	2,24	0,69	1,33
6.	15,00	0,12	1,79	0,74	1,49
7.	19,80	0,08	1,70	0,82	1,53
8.	21,10	0,07	1,51	0,83	1,62
9.	21,10	0,09	1,97	0,72	1,42
10.	18,50	0,10	1,96	0,75	1,42
11.	16,90	0,08	1,50	0,90	1,63
12.	19,01	0,08	1,65	0,84	1,55

Szeged, limit 20 mm

Month	Monthly average mm	Lambda	K	C_v	C_s
1.	26,30	0,08	2,27	0,71	1,32
2.	27,20	0,09	2,68	0,66	1,22
3.	22,70	0,14	3,28	0,56	1,10
4.	17,10	0,14	2,56	0,66	1,24
5.	14,50	0,18	2,62	0,62	1,23
6.	15,30	0,12	1,80	0,73	1,49
7.	19,70	0,10	2,02	0,72	1,40
8.	20,70	0,09	2,02	0,76	1,40
9.	22,30	0,07	1,74	0,84	1,51
10.	22,10	0,07	1,67	0;83	1,54
11.	19,40	0,09	1,93	0,79	1,48
12.	22,80	0,08	1,82	0,73	1'43

tigation of the functional connection. The mean monthly precipitation amounts of the examined stations — the mean values relate to the processed 100 years (1871—1970) — are contained in reference [2].

The analysis of the related values has shown that the best connection is yielded in the case of $C^{(a)}=20$ mm and $C^{(a)}=30$ mm with the precipitation amount of the month under consideration; in the case of $C^{(a)}=50$ mm and $C^{(a)}=70$ mm with that of the month under consideration and in the case of $C^{(a)}=100$ mm with the month under consideration and the two next following months, i.e. with the mean precipitation amounts of three months. The connection $M=f(K)$ for the given threshold value $C^{(a)}$ is shown by *Fig. 4.* On the horizontal axis the part falling to 1 day of the value K of the investigated period can be read while on the vertical axis the values of M.

So e.g. if we want to determine for an initial time t_0 in March and for the threshold value $C^{(a)}=70$ mm the arithmetical mean value M of a precipitation-saturation time $t^{(a)}$ concerning a station where the many years' average of the precipitation amount of May is 41 mm, and that of April 55 mm, then the value to be considered on the horizontal axis will be $(41+55):61=1,57$ and for that precipitation a mean saturation (accumulation) time of $M=52$ days is required.

For the control of the reliability of the approximative computations we selected from our processings at random the times t_0 for each of two stations, and to the empirical distribution functions of the precipitation-saturation time belonging to the two threshold values $C^{(a)}$ fitted the theoretical distribution functions computed with the parameters C_v, C_s, and M taken from the Figures 5 and 6.

The investigated empirical distribution functions are related to the cases

Szeged: t_0=January 1, $C^{(a)}=20$ mm;
Keszthely: t_0=July 15, $C^{(a)}=100$ mm .

After checking the fitting with the computation based on the *Kolmogorov* distribution function the values $p=0{,}11$ and $p=0{,}18$ have been obtained which are satisfying approximations since $p>0{,}05$ and thus the origin of the compared two distributions from an identical set should not be rejected.

References

[1] *Péczely, G. — Herendi, I.:* Csapadéktelítődési idők statisztikai vizsgálata. (Statistical investigation of the times of precipitation saturation) Időjárás, 1976. 93—102. pp.

[2] Magyarország éghajlati atlasza (Climatological Atlas of Hungary) II. kötet, adattár (Vol. II. Data-material) Akadémiai kiadó, Budapest, 1968.

[3] *Yule G. U. — Kendall M. G.:* Introduction into the theory of statistics (In Hungarian translation published by: Közgazdasági és Jogi Könyvkiadó, Budapest, 1964)

[4] *Szigyártó, Z.:* Hidrológiai események valószínűségének becslése eloszlásfüggvények segítségével, (The estimation of the probability of hydrological processes with the aid of distribution functions.) Hidrológiai Közlöny. (Journal of Hydrology)

[5] *Wald, A., Wolfowitz, J.:* An Exact Test for Randomness in the Non-Parametric Case Based on Serial Correlation The Annals of Mathematical Statistics, XIV, v. 1943. pp. 378—388.

[6] *Németh, E.:* Hidrológia és hidrometria (Hydrology and hydrometry) Tankönyvkiadó, Budapest, 1954.